

University of Groningen

Human Lung Tissue Transcriptome

Dugo, Matteo; Cotroneo, Chiara E.; Lavoie-Charland, Emilie; Incarbone, Matteo; Santambrogio, Luigi; Rosso, Lorenzo; van den Berge, Maarten; Nickle, David; Pare, Peter D.; Bosse, Yohan

Published in:
 PLoS ONE

DOI:
 [10.1371/journal.pone.0167460](https://doi.org/10.1371/journal.pone.0167460)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Dugo, M., Cotroneo, C. E., Lavoie-Charland, E., Incarbone, M., Santambrogio, L., Rosso, L., van den Berge, M., Nickle, D., Pare, P. D., Bosse, Y., Dragani, T. A., & Colombo, F. (2016). Human Lung Tissue Transcriptome: Influence of Sex and Age. *PLoS ONE*, 11(11), [e0167460].
 <https://doi.org/10.1371/journal.pone.0167460>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

RESEARCH ARTICLE

Human Lung Tissue Transcriptome: Influence of Sex and Age

Matteo Dugo^{1†}, Chiara E. Cotroneo^{2‡}, Emilie Lavoie-Charland³, Matteo Incarbone⁴, Luigi Santambrogio⁵, Lorenzo Rosso⁵, Maarten van den Berge⁶, David Nickle⁷, Peter D. Paré^{8,9}, Yohan Bossé^{3,10}, Tommaso A. Dragani^{2*}, Francesca Colombo²

1 Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, **2** Department of Predictive and Preventive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, **3** Institut Universitaire de cardiologie et de pneumologie de Québec, Québec, Canada, **4** Department of Surgery, San Giuseppe Hospital–MultiMedica, Milan, Italy, **5** Fondazione IRCCS Ospedale Maggiore Policlinico, University of Milan, Milan, Italy, **6** University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, The Netherlands, **7** Merck & Co. Inc., Rahway, NJ, United States of America, **8** University of British Columbia Center for Heart Lung Innovation and Institute for Heart and Lung Health, St. Paul's Hospital, Vancouver, BC, Canada, **9** Respiratory Division, Department of Medicine, The University of British Columbia, Vancouver, BC, Canada, **10** Department of Molecular Medicine, Laval University, Québec, Canada

‡ Current address: UCD School of Biomolecular and Biomedical Science, University College Dublin, Belfield, Dublin, Ireland.

† These authors are co-first authors on this work.

* tommaso.dragani@istitutotumori.mi.it



OPEN ACCESS

Citation: Dugo M, Cotroneo CE, Lavoie-Charland E, Incarbone M, Santambrogio L, Rosso L, et al. (2016) Human Lung Tissue Transcriptome: Influence of Sex and Age. PLoS ONE 11(11): e0167460. doi:10.1371/journal.pone.0167460

Editor: Baochuan Lin, Defense Threat Reduction Agency, UNITED STATES

Received: September 23, 2015

Accepted: November 15, 2016

Published: November 30, 2016

Copyright: © 2016 Dugo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Gene expression data were deposited in the Gene Expression Omnibus database (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) with accession number GSE71181.

Funding: This work was supported by grants from the Italian Association for Cancer Research (AIRC, grant no. 14714). One of the authors (D.N.) has an affiliation to a commercial funder of this research (Merck & Co. Inc.); the funder provided support in the form of salary for the author (D.N.), but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of

Abstract

Background

Sex and age strongly influence the pathophysiology of human lungs, but scarce information is available about their effects on pulmonary gene expression.

Methods

We followed a discovery-validation strategy to identify sex- and age-related transcriptional differences in lung.

Results

We identified transcriptional profiles significantly associated with sex (215 genes; FDR < 0.05) and age at surgery (217 genes) in non-involved lung tissue resected from 284 lung adenocarcinoma patients. When these profiles were tested in three independent series of non-tumor lung tissue from an additional 1,111 patients, we validated the association with sex and age for 25 and 22 genes, respectively. Among the 17 sex-biased genes mapping on chromosome X, 16 have been reported to escape X-chromosome inactivation in other tissues or cells, suggesting that this mechanism influences lung transcription too. Our 22 age-related genes partially overlap with genes modulated by age in other tissues, suggesting that the aging process has similar consequences on gene expression in different organs. Finally, seven genes whose expression was modulated by sex in non-tumor lung tissue, but no age-related gene, were also validated using publicly available data from 990 lung

this author are articulated in the 'author contributions' section. The funders had no role in the design or conduction of the study, in the collection, analysis or interpretation of the data, or in the preparation, review and approval of the manuscript. The collection of lung samples from Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research (MOP – 123369), and the Cancer Research Society and Read for the Cure. Y. Bossé is the recipient of a Junior 2 Research Scholar award from the Fonds de recherche Québec – Santé (FRQS). The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance.

Competing Interests: The authors declare that they have no competing interests. The affiliation of D.N. to a commercial founder (Merck & Co. Inc.) does not alter our adherence to PLOS ONE policies on sharing data and materials.

adenocarcinoma samples, suggesting that the physiological regulatory mechanisms are only partially active in neoplastic tissue.

Conclusions

Gene expression in non-tumor lung tissue is modulated by both sex and age. These findings represent a validated starting point for research on the molecular mechanisms underlying the observed differences in the course of lung diseases among men and women of different ages.

Introduction

In humans, males and females share a common genome, except for a relatively small number of genes on the Y chromosome; however, the two sexes are noticeably different in morphology, physiology, and behavior. Whole genome analysis of gene expression in different tissues has shown widespread sex differences in the transcriptional profiles of genes [1–3]. This phenomenon, called sex-biased gene expression, has been attributed to the existence of pleiotropic effects of sex on the regulation of gene expression [1–3]. A study in *Drosophila* has indeed shown that female-biased genes on chromosome X have pleiotropic effects [4].

In addition to sex, other biological factors that strongly influence gene expression are age and the process of aging. Indeed, in studies that used genome-wide analyses, age-related changes in transcriptome profiles have been observed in several human tissues, including kidney, muscle, brain, skin, blood and adipose tissue [5–8]. However, there is still open debate on whether the aging process causes similar transcriptional changes in all tissues, and, therefore, if it is possible to identify a common aging signature, or if it predominantly induces tissue-specific molecular changes. Indeed, one study found that only a few genes are commonly affected by age in skin, adipose and brain [6], whereas another found six common age-modulated pathways in a comparison of muscle, kidney and brain [7]. Finally, a meta-analysis of multiple tissues from humans, rats and mice identified common signatures of aging involving, for example, immune/inflammatory responses, cell growth, energy metabolism, and extracellular matrix components [8].

In lung, the effects of age and sex on gene expression are just beginning to be investigated. One study of post-mortem lung tissue from multiorgan donors found 40 genes that were differentially expressed between 7 young and 6 old persons and obtained preliminary results about sex-related differences in gene expression [9]. Substantially more information is available about the effects of age and sex on lung function and cellular physiology. In general, pulmonary function declines with age, even in the absence of respiratory disease. The elderly population has decreased lung volumes, less efficient functionality of respiratory muscles and reduced forced expiratory volumes, accompanied by modifications in innate and adaptive immunity (reviewed in [10]). Lung anatomy and physiology as well as the etiology of some respiratory diseases (e.g. asthma, allergic rhinitis, pulmonary hypertension, cystic fibrosis) have been reported to show sex differences, from prenatal lung development to adulthood; some studies suggested a possible role of sex hormones to explain these differences, but socio-cultural and genetic differences between the sexes are also thought to be involved (reviewed in [11]).

In lung cancer, there are conflicting data on the effects of age and sex. Two large cohort studies reported better survival in women than men [12,13]. Age at diagnosis did not associate with survival in surgically treated patients with stage I non-small cell lung cancer [14]. We

previously observed that neither age at diagnosis nor sex was significantly associated with survival in patients with lung adenocarcinoma [15]. In light of this uncertainty, and considering the relevance of age and sex on lung function in health and disease, greater knowledge about the genes modulated by these biological parameters in lung tissue is needed. Such information is crucial for understanding the molecular mechanisms underlying the observed physiological differences among individuals in the absence of disease, and may also help uncover new targets for the treatment of lung pathologies.

In the present work, we studied the role of sex and age on gene expression in lung tissue, by doing a statistical reanalysis of existing microarray data. We first identified genes affected by sex and age in samples of non-involved lung tissue resected from men and women with lung adenocarcinoma over a broad range of ages (discovery series, [15]). We then determined which of these genes were also differentially expressed in three other series of non-tumor lung tissue (validation series, [16]). Finally, we examined public datasets of gene expression in lung tumors to determine if the sex- and age-biased transcriptomic profiles observed in non-tumor lung tissue are also detectable in lung cancer tissue.

Materials and Methods

Ethics statement

The present study used existing microarray data from four independent clinical series, already described [15,16]. Those papers detailed the approval of the study protocols by the institutional ethics committees (Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, Istituto Clinico Humanitas, Rozzano, Italy, Ospedale San Giuseppe, Milan, Italy, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy, Institut Universitaire de cardiologie et de pneumologie de Québec, Québec, Canada, University of British Columbia, Vancouver, BC, Canada, University of Groningen, Groningen, The Netherlands) and the collection of surgical samples and clinical data from patients who had provided broad informed consent allowing the materials to be used for research purposes; even though follow-up studies were not explicitly authorized, neither were they explicitly refused, and it was expected by all parties involved that the samples and data would support multiple studies. At the recruitment, the authors got the data that have been, then, used in this study. The current analysis did not require additional ethical approval, because it is a continuation of the originally approved studies on data that were full available prior to their use in this study.

Discovery series

The discovery series consisted of 284 samples of non-involved (apparently normal) lung parenchyma (sampled as far as possible from tumor tissue), and the associated clinical data, from patients who had had lobectomy for lung adenocarcinoma at the Fondazione IRCCS Istituto Nazionale dei Tumori and at other hospitals in the area of Milan, Italy. Details about sample collection, RNA extraction and gene expression profiling on Illumina HumanHT-12 v4 Expression BeadChips have already been reported [15]. Because that study also had a discovery-validation design, the microarray analysis had been done on consecutive sets of 206 and 78 samples (although 2 samples from the first set had been excluded from analysis due to the lack of follow-up data).

The analysis of raw microarray data was done at Fondazione IRCCS Istituto Nazionale dei Tumori in Milan essentially as previously described [15]. Briefly, the two datasets of 206 and 78 samples were processed independently using \log_2 transformation and robust spline normalization, implemented in the *lumi* package [17] of the open source software Bioconductor [18]. Then, the two sets were combined using the ComBat adjustment method [19] implemented in

the *sva* R package [20]. Probes that were not annotated and those with a detection *P* value < 0.01 in fewer than 10% of samples were filtered out. Finally, when multiple probes mapped to the same transcript, we included only the one with the highest detection rate, defined as the percentage of samples in which the probe had detection *P* values < 0.01 . Gene expression data were deposited in the Gene Expression Omnibus database (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) with accession number GSE71181.

Validation series

This study also used three validation series, consisting of the discovery and replication sets of another previous work [16]. Here, those sets are called the Laval ($n = 409$), University of British-Columbia (UBC; $n = 339$) and Groningen ($n = 363$) series according to the names of the sites where the patients had lung resection surgery, namely Laval University (Quebec, Canada), University of British-Columbia (Vancouver, Canada), and University of Groningen (Groningen, The Netherlands), respectively. In all cases, non-tumor lung tissues were sampled.

The previous publications [16,21] detailed the sample collection, RNA extraction and gene expression profiling on custom Affymetrix arrays for all three series. Briefly, expression values were extracted from microarray data using the robust multichip average (RMA) method [22]. When two or more probe sets mapped to the same transcript, the one with the greatest number of present calls was selected. The resulting microarray data were deposited in the GEO database with accession number GSE23546.

Datasets on gene expression in lung cancer

To further validate our findings we downloaded, from the GEO database, five public datasets (GSE68896, GSE30219, GSE31210, GSE37745, and GSE41271). GSE68896 contains normalized gene expression data from fetal lung tissue, whereas GSE30219, GSE31210, GSE37745, and GSE41271 contain normalized gene expression data from patients with lung cancer for whom data about age at surgery and sex were available (S1 Table). From each lung cancer dataset, we discarded the records about normal lung tissue samples.

Statistical analysis

Associations of gene expression data with sex and age at surgery (used as dichotomous and continuous variables, respectively) were tested using linear regression modeling. Smoking habit was used as a covariate in the validation series but not in the discovery series, as all samples except one were from smokers, nor in the public lung cancer tissue datasets which lacked information about smoking status.

In the validation phase, the two sets of sex- and age-associated genes were tested independently in each validation series. In both discovery and validation phase, all *P*-values were corrected for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) method [23]; the threshold for statistical significance was set at $\text{FDR} < 0.05$.

To identify sex- and age-biased known biological pathways, we ran gene set enrichment analysis (GSEA) using GSEA v2.0.13 software [24]. Results were visualized using Cytoscape v2.8.3 software [25] and the Enrichment Map plugin [26].

Results

To study the effects of sex and age on gene expression in lung tissue, we analyzed a discovery series of 284 surgical samples from Italy [15] and three validation series from Canada and the Netherlands [16] (Table 1). The discovery series had a predominance of men (71.1%) and all

Table 1. Clinical characteristics of patients from whom a sample of resected non-tumor lung tissue was included in the discovery and validation series.

	Discovery Series (n = 284)	Validation series		
		Laval (n = 409)	UBC (n = 339)	Groningen (n = 363)
Gender				
Men	202 (71.1)	229 (56.0)	182 (53.7)	193 (53.2)
Women	82 (28.9)	180 (44.0)	157 (46.3)	170 (46.8)
Age at surgery, years, mean (SD)	64.7 (9.2)	63.3 (9.9)	60.2 (14.3)	51.5 (15.5)
Smoking status				
Current smoker	283 (99.7) ^a	90 (22.0)	98 (28.9)	57 (15.7)
Former smoker	- ^a	283 (69.2)	163 (48.1)	185 (51.0)
Never-smoker	1 (0.3)	36 (8.8)	26 (7.7)	100 (27.5)
Unknown	0 (0)	0 (0)	52 (15.3)	21 (5.8)
Reason for surgery				
Lung cancer	284 (100)	408 (99.8)	274 (80.8)	123 (33.9)
Other	0 (0)	1 (0.2) ^b	65 (19.2) ^b	240 (66.1) ^b
Lung cancer histotype				
Adenocarcinoma	284 (100)	235 (57.6)	88 (32.1)	37 (30.1)
Squamous cell carcinoma	0 (0)	104 (25.5)	86 (31.4)	54 (43.9)
Carcinoid	0 (0)	17 (4.2)	15 (5.5)	2 (1.6)
Small cell lung carcinoma	0 (0)	7 (1.7)	23 (8.4)	2 (1.6)
Other	0 (0)	45 (11.0)	62 (22.6)	28 (22.8)

Values are n (%) unless indicated otherwise.

^a In the discovery series, individuals are categorized as either never-smokers or ever-smokers, without distinguishing the latter into current smokers or ex-smokers.

^b Patients underwent therapeutic lung resection for lung transplantation or for small peripheral lung lesions.

doi:10.1371/journal.pone.0167460.t001

but one were ever-smokers. The validation series had a more balanced sex distribution, with men representing 53%–56%, and included patients who were never-smokers as well as current and former smokers. In the discovery series, age of both male (Fig 1A, left) and female (Fig 1A, right) patients showed a wide distribution, ranging from 36 to 85 years in males (median = 67 years) and from 40 to 83 years in females (median = 62 years). Age also showed a wide distribution in the three validation sets. For Laval, age ranged from 30 to 82 years in males (Fig 1B, left, median = 66 years) and from 33 to 84 years in females (Fig 1B, right, median = 62 years). For UBC, age ranged from 11 to 85 years in males (Fig 1C, left, median = 64 years) and from 4 to 82 years in females (Fig 1C, right, median = 61 years). For Groningen, age ranged from 6 to 83 years in males (Fig 1D, left, median = 56 years) and from 8 to 75 years in females (Fig 1D, right, median = 53 years).

All patients in the discovery series had been treated for lung adenocarcinoma, while the validation series included patients with a variety of different lung cancer types and even patients who had other lung diseases. In all cancer cases, the surgical samples were apparently normal, having been taken from the non-diseased margins of the resected tissue. Considering the reasons for which the patients had lung resection, their smoking habits and their ages at surgery, the Laval validation series was the most similar to the discovery series.

Discovery of transcripts in lung tissue that associate with sex and age

Expression data were available regarding 11,089 genes for the discovery series [15]. Of these, 215 genes (2%) were differentially expressed between men and women (S2 Table). In

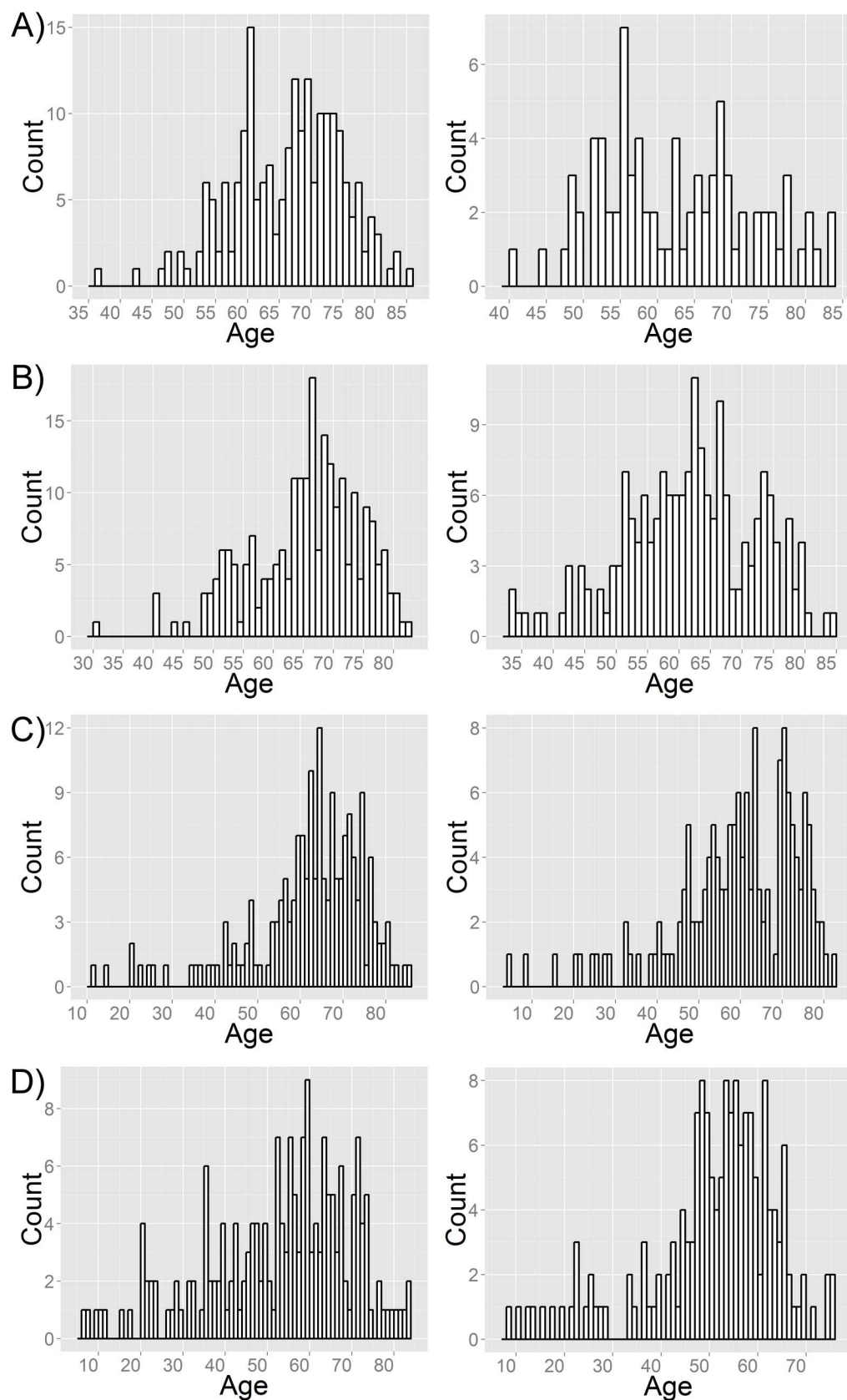


Fig 1. Age distributions of discovery and validation series. Histogram of age at surgery of male (left panels) and female (right panels) lung cancer patients of the discovery (A) and validation series (B: Laval University; C: University of British Columbia; D: University of Groningen).

doi:10.1371/journal.pone.0167460.g001

particular, 122 genes were up-regulated and 93 genes were down-regulated in men compared to women ($FDR < 0.05$). All genes had a fold change between -2 and 2 except for two: *RPS4Y1*, mapping on chromosome Y, was 48-fold up-regulated in men (compared to the background probe intensity measured in females, who do not carry the gene), while *XIST*, mapping on chromosome X, was 12-fold down-regulated. Altogether, 30 sex-associated genes mapped to the non-pseudoautosomal region of chromosome X, one mapped to the non-pseudoautosomal region of chromosome Y, four mapped to the pseudoautosomal regions, and the remaining 180 genes mapped to autosomes. To determine if these genes distinguishing the normal lung tissue of male and female individuals were associated with particular biological pathways or functions, we performed gene set enrichment analysis (GSEA) [24] using gene sets from the Gene Ontology collection (C5.all.v4.0, retrieved from <http://www.broadinstitute.org/gsea/msigdb/index.jsp>). Despite the large number of genes affected by sex, no gene set was found to be enriched in male vs. female individuals.

We then examined the effect of age at surgery on gene expression and found 217 genes that were significantly associated with patients' age ($FDR < 0.05$; [S3 Table](#)). In particular, the expression of 124 genes was positively associated with age, while 93 genes showed a decreased expression in older patients. GSEA revealed that two major functional themes displayed an altered expression in older patients ([Fig 2](#)). The first included gene sets related to extracellular matrix components and functions, with genes encoding, for example, collagens, laminins and metallo-proteinases. The second comprised gene sets related to pro-inflammatory responses and wound healing.

In order to evaluate a possible effect of chance in finding a number of sex- or age-associated genes similar to those observed in this study, we performed a permutation analysis on the discovery series by randomly assigning sex and age labels for 1,000 times; at each permutation cycle, we calculated: i) the total number of genes with $FDR < 0.05$, to determine the expected number of significant genes occurring by chance; ii) the number of times each of the 215 sex-related and 217 age-related genes had an $FDR < 0.05$. We found that in 969 out of 1,000 permutations no genes were significantly associated with sex, with an expected number of genes by chance of 0.721. Results were similar for age, with 972 of 1,000 times with no significant genes and with an expected number of genes by chance of 1.817. This analysis clearly indicates that the limited number of genes that came out of the analysis was not casually generated.

Validation of the sex-biased transcriptional profile

We then attempted to validate the results from the discovery series in three independent validation series (Laval, $n = 409$; UBC, $n = 339$; and Groningen, $n = 363$) [16]. Of the 215 genes associated with sex in the discovery phase, six were not represented on the platform used to profile the validation datasets. For the remaining 209 genes, the possibility of false-positive results due to the high number of statistical tests carried out was taken into account by correcting the obtained *P*-values for multiple testing with the Benjamini–Hochberg procedure and setting a significance threshold of $FDR < 0.05$. With this analysis, 94 genes in the Laval series showed statistically significant and concordant associations with sex and were therefore validated. In addition, 55 genes in the UBC series and 33 in the Groningen series were also validated. Intersection of the lists of genes significantly associated with sex in each of the four datasets (discovery, Laval, UBC and Groningen) highlighted a common set of 25 genes that were validated in all series ([Fig 3](#)).

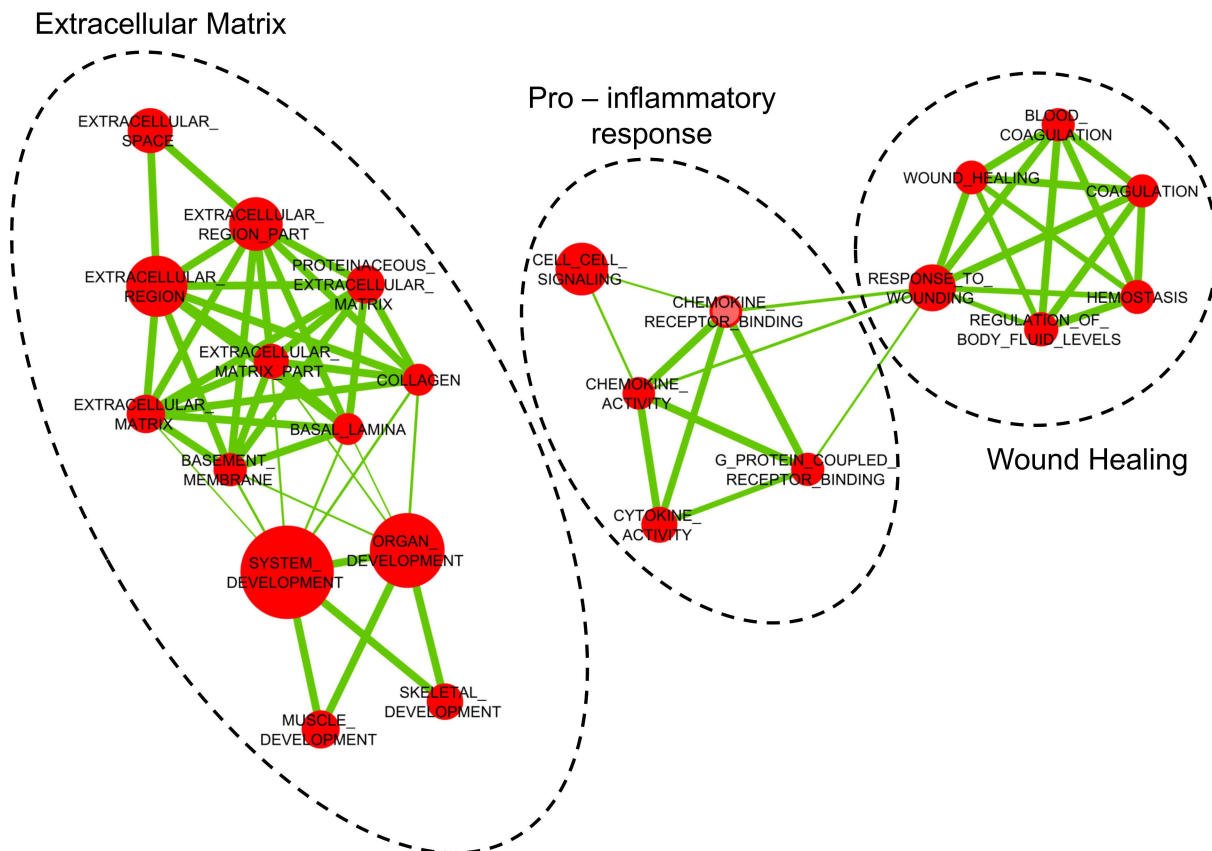


Fig 2. Two major functional themes displayed an altered expression in older patients. Network-based visualization of gene sets enriched in lung tissue in association with age, in 284 patients who underwent lobectomy for lung adenocarcinoma; the analyzed lung tissue was not involved by cancer. Gene set enrichment analysis (GSEA) was carried out using GSEA v2.0.13 software [24]. Briefly, we first ranked genes of the discovery dataset according to the t-statistic of age and sex. Then, gene sets derived from the Gene Ontology database (C5.all.v4.0) were retrieved from the Molecular Signature Database (MSigDB, <http://www.broadinstitute.org/gsea/msigdb/index.jsp>) and tested for enrichment using a Kolmogorov-Smirnov statistic. Only the 653 gene sets having between 15 and 500 genes were tested for enrichment. An FDR < 0.05 identified significantly enriched gene sets. GSEA results were visualized using Cytoscape v2.8.3 software [25] and the Enrichment Map plugin [26], with an overlap coefficient cut-off of 0.5. Circles (nodes) represent C5 Gene Ontology gene sets connected by lines (edges) whose thickness is proportional to the number of genes shared between the connected nodes. Circle sizes are proportional to the number of genes annotated in each gene set. Two independent clusters of functionally related gene sets were detected, one (on the left) involving extracellular matrix and the other one (on the right) involving pro-inflammatory response and wound healing.

doi:10.1371/journal.pone.0167460.g002

This common set of sex-biased genes in non-tumor lung tissue included five genes that were up-regulated and 20 that were down-regulated in men compared to women. Overall, five of the genes mapped to four different autosomes, 17 mapped to the X chromosome and one to the Y chromosome (pseudo-autosomal regions excluded), and two mapped to the pseudoautosomal regions (Table 2). The two genes that showed an absolute fold change greater than 2 between males and females in the discovery series (*RPS4Y1* and *XIST*) were included in this set.

To assess whether transcriptional differences in the lung of males and females arise at early stages of development, we analyzed public gene expression data of fetal lung tissues (GSE68896, downloaded from the Gene Expression Omnibus database at <http://www.ncbi.nlm.nih.gov/geo/>). Interestingly, we found that 18 of the 25 validated sex-related genes identified in adult lung tissue (Table 2) were also differentially expressed between male and female fetal lung tissues (S4 Table).

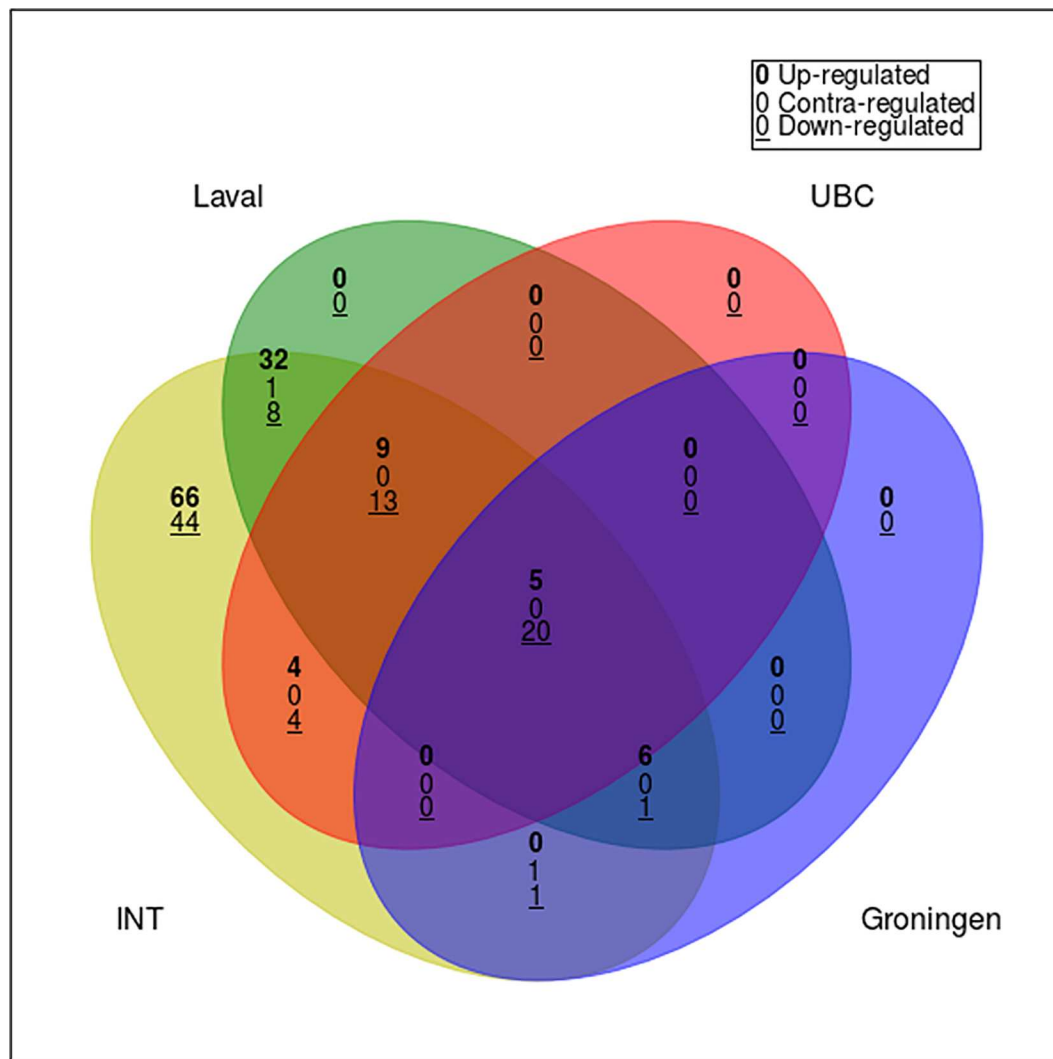


Fig 3. Intersection of the lists of genes significantly associated with sex in each of the four datasets. Four-way Venn diagram analysis of sex-related transcriptional alterations in non-tumor lung tissue. Each of the circles depicts the number of different transcripts based on a sex comparison for each of the labeled data series (yellow, INT, Italian discovery series; green, Laval validation series; blue, Groningen validation series; red, UBC validation series) among the 215 transcripts identified as statistically significant in the Italian discovery series. Shared transcripts are represented in the areas of intersection between two or more circles. Genes whose expression levels were positively and negatively associated with sex in the different series are called “contra-regulated”; these genes were not included among the number of validated genes.

doi:10.1371/journal.pone.0167460.g003

We also attempted to validate the results from the discovery series in four publicly available datasets of gene expression in lung cancer samples (GSE30219, GSE31210, GSE37745, and GSE41271). As these four datasets were profiled with different platforms, data were available for 169 of the 215 genes associated with sex in the discovery series. Among these 169 genes, 52 showed a statistically significant ($FDR < 0.05$) and concordant association with sex in at least one cancer dataset (not shown). The discovery series and the four tumor series shared a common set of seven sex-biased genes, namely *ARSD*, *CD99*, *GEMIN8*, *OFD1*, *RPS4X*, *RPS4Y1*, and *XIST*. Interestingly, all of these genes also belonged to the common set of genes validated in all three series of non-tumor lung tissue.

Validation of the age-biased transcriptional profile

Among the 217 genes associated with age at surgery in the discovery series, 13 had not been investigated in the three validation series (Laval, UBC, Groningen). After correcting for multiple testing, the association with age (with the same direction of effect) was confirmed for 149 genes in the Laval series, 97 genes in the UBC series, and 32 genes in the Groningen series ($FDR < 0.05$). The intersection of these sets identified 22 common genes showing a concordant direction of effect in all four series (Fig 4). These included 15 up-regulated and 7 down-regulated genes in older individuals, which were found on 14 different autosomes (Table 3).

The same analysis was repeated for the four public datasets of lung cancer specimens (GSE30219, GSE31210, GSE37745, and GSE41271). Of the 217 age-associated genes found in the discovery series, 178 had also been investigated in the four tumor series. Statistical analysis showed that only one gene (*IRX3*) maintained a significant ($FDR < 0.05$) and concordant association in at least one of the four cancer datasets (not shown), and none maintained the association in all four datasets. Therefore, it was not possible to identify a common set of genes regulated by age in lung tumor tissue.

Discussion

In our discovery series of 284 lung adenocarcinoma patients, we identified 215 and 217 genes whose expression levels in non-involved lung tissue were significantly associated with sex and

Table 2. Genes whose expression in lung tissue associated with sex in the discovery series and in all three validation series.

Gene symbol	Gene name	Gene Entrez ID	Chr.	Fold change ¹	FDR ¹
<i>AOX1</i>	Aldehyde oxidase 1	316	2	1.397	2.81E-04
<i>ARSD</i>	Arylsulfatase D	414	X	-1.543	5.78E-50
<i>ATP6V1B1</i>	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B1	525	2	-1.352	3.30E-04
<i>CD99</i>	CD99 molecule	4267	X;Y	1.342	5.85E-22
<i>CILP</i>	Cartilage intermediate layer protein, nucleotide pyrophosphohydrolase	8483	15	1.397	1.24E-02
<i>DDX3X</i>	DEAD (Asp-Glu-Ala-Asp) box helicase 3, X-linked	1654	X	-1.215	2.29E-04
<i>EIF1AX</i>	Eukaryotic translation initiation factor 1A, X-linked	1964	X	-1.336	1.34E-16
<i>EIF2S3</i>	Eukaryotic translation initiation factor 2, subunit 3 gamma, 52kDa	1968	X	-1.231	1.39E-16
<i>GEMIN8</i>	Gem (nuclear organelle) associated protein 8	54960	X	-1.134	1.20E-04
<i>HDHD1</i>	Haloacid dehalogenase-like hydrolase domain containing 1	8226	X	-1.586	1.75E-50
<i>KAL1</i>	Kallmann syndrome 1 sequence	3730	X	-1.982	1.75E-30
<i>KDM6A</i>	Lysine (K)-specific demethylase 6A	7403	X	-1.514	4.96E-39
<i>OFD1</i>	Oral-facial-digital syndrome 1	8481	X	-1.179	5.06E-04
<i>OLFML2A</i>	Olfactomedin-like 2A	169611	9	-1.287	6.86E-03
<i>PHGDH</i>	Phosphoglycerate dehydrogenase	26227	1	-1.170	3.64E-02
<i>PRKX</i>	Protein kinase, X-linked	5613	X	-1.131	2.38E-03
<i>RPS4X</i>	Ribosomal protein S4, X-linked	6191	X	-1.369	1.83E-09
<i>RPS4Y1</i>	Ribosomal protein S4, Y-linked 1	6192	Y	48.948	1.76E-249
<i>TRAPPC2</i>	Trafficking protein particle complex 2	6399	X	-1.171	1.72E-07
<i>TXLNG</i>	Taxilin gamma	55787	X	-1.184	7.33E-07
<i>USP9X</i>	Ubiquitin specific peptidase 9, X-linked	8239	X	-1.084	2.15E-02
<i>XIST</i>	X inactive specific transcript (non-protein coding)	7503	X	-12.561	1.14E-171
<i>ZBED1</i>	Zinc finger, BED-type containing 1	9189	X;Y	1.108	3.91E-05
<i>ZFX</i>	Zinc finger protein, X-linked	7543	X	-1.382	3.14E-21
<i>ZRSR2</i>	Zinc finger (CCCH type), RNA-binding motif and serine/arginine rich 2	8233	X	-1.362	2.54E-16

¹ Values are those observed in the discovery series; males versus females, values > 1 indicate higher levels in males.

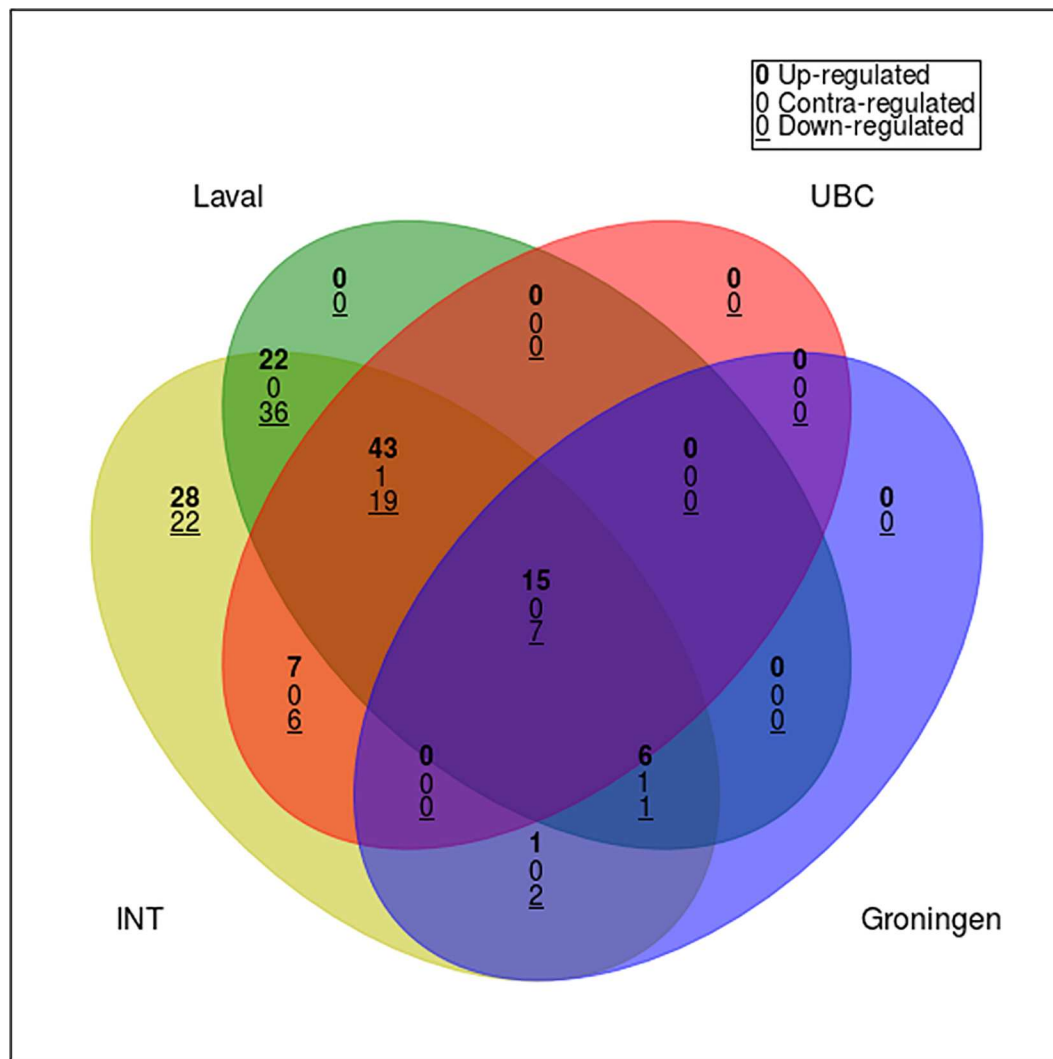


Fig 4. Intersection of the lists of genes significantly associated with age in each of the four datasets. Four-way Venn diagram analysis of age-related transcriptional alterations in non-tumor lung tissue. As in Fig 3, each of the circles depicts the number of different transcripts based on age comparison for each of the labeled data series (yellow, INT, Italian discovery series; green, Laval validation series; blue, Groningen validation series; red, UBC validation series) among the 217 transcripts identified as statistically significant in the Italian discovery series. Shared transcripts are represented in the areas of intersection between two or more circles. Genes whose expression levels were positively and negatively associated with age in the different series are called “contra-regulated”; these genes were not included among the number of validated genes.

doi:10.1371/journal.pone.0167460.g004

age, respectively. GSEA analysis revealed that no specific molecular functions were over-represented among the genes altered in males vs. females, whereas genes whose expression was altered in older patients belonged to two major functional categories, i.e. extracellular matrix function and pro-inflammatory response/wound healing. We validated 25 out of 215 sex-related genes and 22 out of 217 age-related genes in three independent series of non-tumor lung tissue (from a total of 1,111 individuals). Additionally, we validated a common set of seven sex-biased genes in four publicly available datasets of transcriptional profiles of lung tumor tissue (990 total individuals); no common age-associated genes were identified in all four lung cancer datasets, and only one gene was validated in at least one of the four tumor series.

Table 3. Genes whose expression in tissue associated with patients' age at surgery in the discovery series and in all three validation series.

Symbol	Gene name	Gene Entrez ID	Chr.	Estimate ¹	FDR ¹
<i>AGR2</i>	Anterior gradient 2	10551	7	-0.01213	2.95E-02
<i>CRYAB</i>	Crystallin, alpha B	1410	11	0.013824	1.60E-03
<i>CXCL17</i>	Chemokine (C-X-C motif) ligand 17	284340	19	-0.00911	4.85E-02
<i>CXCL9</i>	Chemokine (C-X-C motif) ligand 9	4283	4	0.028698	3.35E-02
<i>DIRAS3</i>	DIRAS family, GTP-binding RAS-like 3	9077	1	0.008936	3.00E-03
<i>FANCE</i>	Fanconi anemia, complementation group E	2178	6	-0.01075	1.10E-03
<i>FMO3</i>	Flavin containing monooxygenase 3	2328	1	0.0131	1.44E-02
<i>GSTA4</i>	Glutathione S-transferase A4	2941	6	-0.01173	4.80E-03
<i>ITGB5</i>	Integrin, beta 5	3693	3	0.007076	4.19E-02
<i>ITGBL1</i>	Integrin, beta-like 1 (with EGF-like repeat domains)	9358	13	0.018297	1.21E-05
<i>LCN2</i>	Lipocalin 2 (oncogene 24p3)	3934	9	-0.02576	3.78E-02
<i>MGP</i>	Matrix Gla protein	4256	12	0.00841	4.97E-02
<i>PGM5</i>	Phosphoglucomutase 5	5239	9	0.009683	4.80E-02
<i>PLCD3</i>	Phospholipase C, delta 3	113026	17	0.005974	1.42E-02
<i>PTCHD4</i>	Patched domain containing 4	442213	6	0.003035	4.85E-02
<i>RCAN2</i>	Regulator of calcineurin 2	10231	6	0.015413	1.56E-04
<i>RERGL</i>	RERG/RAS-like	79785	12	0.022899	6.70E-03
<i>SCARF2</i>	Scavenger receptor class F, member 2	91179	22	0.009692	1.96E-02
<i>TC2N</i>	Tandem C2 domains, nuclear	123036	14	-0.00923	3.78E-02
<i>WISP2</i>	WNT1 inducible signaling pathway protein 2	8839	20	0.026326	9.02E-05
<i>ZMAT3</i>	Zinc finger, matrin type 3	64393	3	0.016636	1.84E-04
<i>ZNF518B</i>	Zinc finger protein 518B	85460	4	-0.00626	4.07E-02

¹ Values are those observed in the discovery series.

doi:10.1371/journal.pone.0167460.t003

Regarding the effect of sex on lung tissue transcription, we observed that 17 of the 25 sex-biased genes map to the non-pseudoautosomal region of chromosome X and displayed higher expression in females. These genes have probably escaped X-chromosome inactivation, i.e. the mechanism by which mammals balance gene expression between the sexes through the silencing of one X chromosome in somatic cells of females [27]. Indeed, all these genes have been observed to escape X-chromosome inactivation in an *in vitro* assay using rodent/human somatic cell hybrids [28], with the exception of *ZRSR2* which was not tested in that study. X-chromosome inactivation is regulated by several factors, including the *XIST* gene which we found to be highly up-regulated (12 fold) in females. Furthermore, four genes identified here (*HDHD1*, *RPS4X*, *KDM6A*, and *ZFX*) were found to have higher expression in females in various human tissues [29], although in that study only *KDM6A* (alias *UTX*) was differentially expressed in lung tissue. We also observed the sex-biased expression of genes never reported to escape from X-inactivation in human tissue and also of genes mapping on autosomes, suggesting the existence of other mechanisms of sex-dependent transcriptional control in lung tissue. As we did not find any functional category enriched among male- or female-specific genes, we cannot hypothesize about the mechanisms involved. Of note, in our analysis we did not consider the possible confounding effects of hormonal status on gene expression. Due to the small number ($n = 8$ women younger than 50 years of age) of non- or pre-menopausal females in our series, we doubt that menopausal status had any relevant impact on our results. Indeed, any subgroup analysis based on age in females would be severely biased by a very small sample size in one subgroup. Finally, we also noticed that $>70\%$ (i.e., 18/25) of the sex-related genes that passed validation showed a differential expression in males and females

already in fetal lung, indicating that sexual differences in lung transcriptome are, at least in part, established at the developmental level.

Regarding age-dependent modulation of transcription, our study builds on preliminary evidence [9] of age-related changes in gene expression in non-tumor lung tissue and defines a lung-specific aging signature of 22 genes. Only the *HEPH* gene of our 22 genes was also present in the preliminary age-related signature of 40 genes reported in [9]. Five genes in our signature have previously been associated with aging in skin and/or adipose tissue [6]. In particular, two of them (*DIRAS3*, and *WISP2*) were found to be up-regulated in aged skin; two genes (*CXCL9* and *RCAN2*) were up-regulated with age in adipose tissue; and *FMO3* was reported up-regulated in both aged skin and adipose tissue. Their involvement in aging lung, reported here, suggests that these genes may have broad roles in the aging process.

None of the 22 age-related genes identified in this study belongs to the transcriptional signature of aging defined in a meta-analysis of multiple tissues from humans, rat and mice [8]. That study identified age-dependent modifications in the expression of genes involved in several biological processes, including pro-inflammatory response genes, and extracellular matrix function. We, too, found age-related changes in the expression of genes involved in pro-inflammatory responses and genes encoding extracellular matrix components, in our discovery series. Although there were no common genes between our study and the meta-analysis described in [8], pathways related to the pro-inflammatory responses and to the extracellular matrix are important in lung function and disease. The known age-related increase in pulmonary inflammation, as observed for example in lungs of patients with chronic obstructive pulmonary disease (COPD) [30], may reflect the lungs' exposure, over the course of a lifetime, to environmental pollutants and microorganisms. Also, age-related changes in extracellular matrix alter the mechanisms of lung repair and lead to abnormal wound healing and fibrosis (reviewed in [31]). Such alterations are often observed, respectively, in COPD and in idiopathic pulmonary fibrosis, both associated with aging [32,33]. Therefore, the age-related genes identified in our study are worthy of further investigation to understand their possible involvement in these pulmonary diseases.

Interestingly, seven genes belonging to the sex-associated profile from non-involved lung tissue were also validated in four lung cancer series. This finding suggests that some sex-related transcriptome changes are maintained in neoplastic tissue which, therefore, is controlled to a certain degree by the same sex-linked factors active in normal lung tissue. In the same four lung cancer series, however, we did not validate any genes from the age-related expression profile and only one gene showed a significant association with age in at least one of the four datasets. This result suggests that gene expression in lung tumor tissue is not affected by factors associated with aging in non-tumor tissue.

One limitation of our study is that we do not have more details about patients' smoking habits (e.g. years of smoking, packs per year) in the discovery series; thus we could not take into account this aspect in our analyses. Also, the study was carried out in whole lung tissue; therefore, the cell-specificity of expression of the validated genes should be further investigated by immunohistochemical studies.

Two other limitations of our study, which might explain the low number of validated sex- and age-related genes, are the use of different platforms for gene expression analysis in the discovery versus validation series, and the high degree of heterogeneity, in terms of patients' ages, sex distribution, smoking habits, and types of lung pathology, of the validation series compared with the discovery series. Indeed, the number of validated genes is quite low (around 10% of those found in the discovery series for both sex and age), but the validation was carried out on three independent datasets. Because this approach is stringent, it results in a lower number of validated genes than had we used only a single validation dataset, as seen in the

majority of published reports. The relatively few genes validated in our study represent the core of genes that are always significantly associated with sex or age in all of the datasets analyzed; we believe that these are true positives. If we are less stringent and consider the genes found significant in at least one dataset, the numbers of validated genes increase from 25 to 102 for sex and from 22 to 165 for age.

It is interesting to notice that more genes were validated for sex and age in the Laval series, which is clinically most similar to the Italian discovery series, than in the other two validation series. This finding suggests that the transcriptional profiles identified in the Italian discovery series contain a high percentage of real positive associations. This is also supported by permutation analysis results indicating a very low possibility that the obtained numbers of sex- or age-related genes were due to chance.

The mechanisms underlying the observed differential sex or age expression of genes in lung may be due to differences in DNA methylation. Indeed, sex differences in DNA methylation, leading to differential gene expression at various loci, have been reported [34]. Moreover, age-related changes in DNA methylation have also been observed [35].

Overall, we found that, in non-tumor lung tissue, several genes undergo a modulation of their expression, which depends on either sex or age of the examined individual. Our findings provide a reliable starting point for a deeper investigation at the molecular level of the role of sex and age in the pathophysiology of lung tissue.

Supporting Information

S1 Table. Clinical characteristics of patients with lung cancer whose resected tumor tissue was analyzed for gene expression, according to the GEO dataset in which the data were deposited
(DOCX)

S2 Table. List of 215 genes whose expression in lung tissue associated with gender in the discovery series, in order of false discovery rate (FDR) values.
(DOCX)

S3 Table. List of 217 genes whose expression in lung tissue associated with age at surgery in the discovery series, presented in order of false discovery rate (FDR) values.
(DOCX)

S4 Table. Sex-related differentially expressed genes in both adult and fetal lung tissue. Of the 25 validated sex-related genes identified in adult lung tissue, 18 were also found to be differentially expressed between males and females in fetal lung tissue (GSE68896)
(DOCX)

Acknowledgments

We thank Valerie Matarese for scientific editing of the manuscript.

Author Contributions

Conceptualization: FC MvdB TAD YB.

Formal analysis: CEC ELC MD.

Funding acquisition: TAD DN YB.

Investigation: DN.

Resources: LR LS MI MvdB PDP YB.

Supervision: FC MvdB TAD YB.

Writing – original draft: CEC FC MD PDP TAD YB.

Writing – review & editing: CEC FC MD PDP TAD YB.

References

1. Jansen R, Batista S, Brooks AI, Tischfield JA, Willemsen G, van Grootheest G, et al. Sex differences in the human peripheral blood transcriptome. *BMC Genomics*. 2014; 15: 33-2164-15-33.
2. Xu H, Wang F, Liu Y, Yu Y, Gelernter J, Zhang H. Sex-biased methylome and transcriptome in human prefrontal cortex. *Hum Mol Genet*. 2014; 23: 1260–1270. doi: [10.1093/hmg/ddt516](https://doi.org/10.1093/hmg/ddt516) PMID: [24163133](https://pubmed.ncbi.nlm.nih.gov/24163133/)
3. Ellegren H, Parsch J. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet*. 2007; 8: 689–698. doi: [10.1038/nrg2167](https://doi.org/10.1038/nrg2167) PMID: [17680007](https://pubmed.ncbi.nlm.nih.gov/17680007/)
4. Assis R, Zhou Q, Bachtrög D. Sex-biased transcriptome evolution in *Drosophila*. *Genome Biol Evol*. 2012; 4: 1189–1200. doi: [10.1093/gbe/evs093](https://doi.org/10.1093/gbe/evs093) PMID: [23097318](https://pubmed.ncbi.nlm.nih.gov/23097318/)
5. Rodwell GE, Sonu R, Zahn JM, Lund J, Wilhelmy J, Wang L, et al. A transcriptional profile of aging in the human kidney. *PLoS Biol*. 2004; 2: e427. doi: [10.1371/journal.pbio.0020427](https://doi.org/10.1371/journal.pbio.0020427) PMID: [15562319](https://pubmed.ncbi.nlm.nih.gov/15562319/)
6. Glass D, Vinuela A, Davies MN, Ramasamy A, Parts L, Knowles D, et al. Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol*. 2013; 14: R75. doi: [10.1186/gb-2013-14-7-r75](https://doi.org/10.1186/gb-2013-14-7-r75) PMID: [23889843](https://pubmed.ncbi.nlm.nih.gov/23889843/)
7. Zahn JM, Sonu R, Vogel H, Crane E, Mazan-Mamczarz K, Rabkin R, et al. Transcriptional profiling of aging in human muscle reveals a common aging signature. *PLoS Genet*. 2006; 2: e115. doi: [10.1371/journal.pgen.0020115](https://doi.org/10.1371/journal.pgen.0020115) PMID: [16789832](https://pubmed.ncbi.nlm.nih.gov/16789832/)
8. de Magalhaes JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*. 2009; 25: 875–881. doi: [10.1093/bioinformatics/btp073](https://doi.org/10.1093/bioinformatics/btp073) PMID: [19189975](https://pubmed.ncbi.nlm.nih.gov/19189975/)
9. Gruber MP, Coldren CD, Woolum MD, Cosgrove GP, Zeng C, Baron AE, et al. Human lung project: evaluating variance of gene expression in the human lung. *Am J Respir Cell Mol Biol*. 2006; 35: 65–71. doi: [10.1165/rcmb.2004-0261OC](https://doi.org/10.1165/rcmb.2004-0261OC) PMID: [16498083](https://pubmed.ncbi.nlm.nih.gov/16498083/)
10. Lowery EM, Brubaker AL, Kuhlmann E, Kovacs EJ. The aging lung. *Clin Interv Aging*. 2013; 8: 1489–1496. doi: [10.2147/CIA.S51152](https://doi.org/10.2147/CIA.S51152) PMID: [24235821](https://pubmed.ncbi.nlm.nih.gov/24235821/)
11. Townsend EA, Miller VM, Prakash YS. Sex differences and sex steroids in lung health and disease. *Endocr Rev*. 2012; 33: 1–47. doi: [10.1210/er.2010-0031](https://doi.org/10.1210/er.2010-0031) PMID: [22240244](https://pubmed.ncbi.nlm.nih.gov/22240244/)
12. Tong BC, Kosinski AS, Burfeind WR Jr, Onaitis MW, Berry MF, Harpole DH Jr, et al. Sex differences in early outcomes after lung cancer resection: analysis of the Society of Thoracic Surgeons General Thoracic Database. *J Thorac Cardiovasc Surg*. 2014; 148: 13–18. doi: [10.1016/j.jtcvs.2014.03.012](https://doi.org/10.1016/j.jtcvs.2014.03.012) PMID: [24726742](https://pubmed.ncbi.nlm.nih.gov/24726742/)
13. Sagerup CM, Smastuen M, Johannesen TB, Helland A, Brustugun OT. Sex-specific trends in lung cancer incidence and survival: a population study of 40,118 cases. *Thorax*. 2011; 66: 301–307. doi: [10.1136/thx.2010.151621](https://doi.org/10.1136/thx.2010.151621) PMID: [21199818](https://pubmed.ncbi.nlm.nih.gov/21199818/)
14. Palma DA, Tyldesley S, Sheehan F, Mohamed IG, Smith S, Wai E, et al. Stage I non-small cell lung cancer (NSCLC) in patients aged 75 years and older: does age determine survival after radical treatment? *J Thorac Oncol*. 2010; 5: 818–824. PMID: [20521349](https://pubmed.ncbi.nlm.nih.gov/20521349/)
15. Galvan A, Frullanti E, Anderlini M, Manenti G, Noci S, Dugo M, et al. Gene expression signature of non-involved lung tissue associated with survival in lung adenocarcinoma patients. *Carcinogenesis*. 2013; 34: 2767–2773. doi: [10.1093/carcin/bgt294](https://doi.org/10.1093/carcin/bgt294) PMID: [23978379](https://pubmed.ncbi.nlm.nih.gov/23978379/)
16. Bosse Y, Postma DS, Sin DD, Lamontagne M, Couture C, Gaudreault N, et al. Molecular signature of smoking in human lung tissues. *Cancer Res*. 2012; 72: 3753–3763. doi: [10.1158/0008-5472.CAN-12-1160](https://doi.org/10.1158/0008-5472.CAN-12-1160) PMID: [22659451](https://pubmed.ncbi.nlm.nih.gov/22659451/)
17. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008; 24: 1547–1548. doi: [10.1093/bioinformatics/btn224](https://doi.org/10.1093/bioinformatics/btn224) PMID: [18467348](https://pubmed.ncbi.nlm.nih.gov/18467348/)
18. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5: R80. doi: [10.1186/gb-2004-5-10-r80](https://doi.org/10.1186/gb-2004-5-10-r80) PMID: [15461798](https://pubmed.ncbi.nlm.nih.gov/15461798/)
19. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8: 118–127. doi: [10.1093/biostatistics/kxj037](https://doi.org/10.1093/biostatistics/kxj037) PMID: [16632515](https://pubmed.ncbi.nlm.nih.gov/16632515/)

20. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012; 28: 882–883. doi: [10.1093/bioinformatics/bts034](https://doi.org/10.1093/bioinformatics/bts034) PMID: [22257669](https://pubmed.ncbi.nlm.nih.gov/22257669/)
21. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet*. 2012; 8: e1003029. doi: [10.1371/journal.pgen.1003029](https://doi.org/10.1371/journal.pgen.1003029) PMID: [23209423](https://pubmed.ncbi.nlm.nih.gov/23209423/)
22. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4: 249–264. doi: [10.1093/biostatistics/4.2.249](https://doi.org/10.1093/biostatistics/4.2.249) PMID: [12925520](https://pubmed.ncbi.nlm.nih.gov/12925520/)
23. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B*. 1995; 57: 289–300.
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102: 15545–15550. doi: [10.1073/pnas.0506580102](https://doi.org/10.1073/pnas.0506580102) PMID: [16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13: 2498–2504. doi: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303) PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
26. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*. 2010; 5: e13984. doi: [10.1371/journal.pone.0013984](https://doi.org/10.1371/journal.pone.0013984) PMID: [21085593](https://pubmed.ncbi.nlm.nih.gov/21085593/)
27. Berletch JB, Yang F, Xu J, Carrel L, Distech CM. Genes that escape from X inactivation. *Hum Genet*. 2011; 130: 237–245. doi: [10.1007/s00439-011-1011-z](https://doi.org/10.1007/s00439-011-1011-z) PMID: [21614513](https://pubmed.ncbi.nlm.nih.gov/21614513/)
28. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*. 2005; 434: 400–404. doi: [10.1038/nature03479](https://doi.org/10.1038/nature03479) PMID: [15772666](https://pubmed.ncbi.nlm.nih.gov/15772666/)
29. Talebizadeh Z, Simon SD, Butler MG. X chromosome gene expression in human tissues: male and female comparisons. *Genomics*. 2006; 88: 675–681. doi: [10.1016/j.ygeno.2006.07.016](https://doi.org/10.1016/j.ygeno.2006.07.016) PMID: [16949791](https://pubmed.ncbi.nlm.nih.gov/16949791/)
30. Ito K, Colley T, Mercado N. Geroprotectors as a novel therapeutic strategy for COPD, an accelerating aging disease. *Int J Chron Obstruct Pulmon Dis*. 2012; 7: 641–652. doi: [10.2147/COPD.S28250](https://doi.org/10.2147/COPD.S28250) PMID: [23055713](https://pubmed.ncbi.nlm.nih.gov/23055713/)
31. Kapetanaki MG, Mora AL, Rojas M. Influence of age on wound healing and fibrosis. *J Pathol*. 2013; 229: 310–322. doi: [10.1002/path.4122](https://doi.org/10.1002/path.4122) PMID: [23124998](https://pubmed.ncbi.nlm.nih.gov/23124998/)
32. Papaioannou AI, Rossios C, Kostikas K, Ito K. Can we delay the accelerated lung aging in COPD? Anti-aging molecules and interventions. *Curr Drug Targets*. 2013; 14: 149–157. PMID: [23256715](https://pubmed.ncbi.nlm.nih.gov/23256715/)
33. Renzoni E, Srihari V, Sestini P. Pathogenesis of idiopathic pulmonary fibrosis: review of recent findings. *F1000Prime Rep*. 2014; 6: 69–69. eCollection 2014. doi: [10.12703/P6-69](https://doi.org/10.12703/P6-69) PMID: [25165568](https://pubmed.ncbi.nlm.nih.gov/25165568/)
34. Singmann P, Shem-Tov D, Wahl S, Grallert H, Fiorito G, Shin SY, et al. Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics Chromatin*. 2015; 8: 43-015-0035-3. eCollection 2015.
35. Bacalini MG, Boattini A, Gentilini D, Giampieri E, Pirazzini C, Giuliani C, et al. A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data. *Aging (Albany NY)*. 2015; 7: 97–109.